DTIC FILE COPY

(4)

CRM 87-185 / October 1987

RESEARCH MEMORANDUM

AD-A196 623

BAYESIAN ESTIMATION OF *n* IN A BINOMIAL DISTRIBUTION

Victor K. T. Tang
Ronald B. Sindler
Raymond M. Shirven

A Division of CNA Hudson Institute

CENTER FOR NAVAL ANALYSES

4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

88 6 7 048

Work conducted under contract N00014-87-C-0001.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) CRM 87-185	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Center for Naval Analyses — 6b. OFFICE SYMBOL (If applicable) CNA	7a. NAME OF MONITORING ORGANIZATION Office of the Chief of Naval Operations (OP-91)
6c. ADDRESS (City, State, and ZIP Code) 4401 Ford Avenue Alexandria, Virginia 22302-0268	7b. ADDRESS (City, State, and ZIP Code) Navy Department Washington, D.C. 20350-2000
8a. NAME OF FUNDING / ORGANIZATION Office of Naval Research — 8b. OFFICE SYMBOL (If applicable) ONR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-87-C-0001
8c. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, Virginia 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 65154N PROJECT NO. R0148 TASK NO. WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Bayesian Estimation of n in a Binomial Distribution

12. PERSONAL AUTHOR(S)

Victor K. T. Tang, Ronald B. Sindler, Raymond M. Shirven

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM TO	October 1987	26

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Let X_1, X_2, ..., X_r be a random sample of size r from a binomial distribution $b(n, p)$. Let x_1, x_2, ..., x_r be r-observed success counts. A method has been developed to estimate the total number of trials n from a Bayesian perspective when the probability of success p is either known or unknown. The prior distribution for n is assumed to be the discrete uniform distribution. In the case when p is unknown, p is assumed to have a beta prior distribution. The estimate for p is then the mode of the posterior distribution. Additionally, guidelines for selecting shape parameters for the beta distributions are discussed.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) — 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

CRM 87-185 / October 1987

BAYESIAN ESTIMATION OF *n* IN A BINOMIAL DISTRIBUTION

Victor K. T. Tang
Ronald B. Sindler
Raymond M. Shirven

Strike and Amphibious Warfare Research Department

A Division of CNA *Hudson Institute*

CENTER FOR NAVAL ANALYSES

4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268

ABSTRACT

Let $X_1, X_2, \ldots, X_r$ be a random sample of size r from a binomial distribution $b(n, p)$. Let $x_1, x_2, \ldots, x_r$ be r-observed success counts. A method has been developed to estimate the total number of trials n from a Bayesian perspective when the probability of success p is either known or unknown. The prior distribution for n is assumed to be the discrete uniform distribution. In the case when p is unknown, p is assumed to have a beta prior distribution. The estimate for n is then the mode of the posterior distribution. Additionally, guidelines for selecting shape parameters for the beta distributions are discussed. Keywords: Fortran, [illegible] (KR)

TABLE OF CONTENTS

INTRODUCTION

Consider that $X_1, X_2, \ldots, X_r$ is a random sample of size r from a binomial distribution $b(n, p)$, where n is the number of trials and p is the probability of success. Given r observations, $x_1, x_2, \ldots, x_r$, the usual problem in the binomial situation is to estimate p. However, in other instances, n, the number of trials, may be the unknown parameter of interest. For example, an appliance company may be interested in estimating the number of appliances of a particular type in a given service area. In the naval environment, when a number of ships have been located or detected in a certain region, sometimes it is important to estimate the overall number of ships in that region in order to decide appropriate actions. These are some of the practical problems in which n in a binomial distribution becomes the parameter to be estimated.

In the simplest case, when $r = 1$ and p is known, the procedure for estimating n by constructing confidence intervals has been worked out by Tang and Sindler in [1]. This research is a follow-up work in which the general cases are considered but some of the restrictions are removed. r is assumed to be a positive integer greater than one and p may or may not be known. Draper and Guttman [2] proposed to estimate n in a Bayesian framework by introducing prior probability distributions to unknown parameters. Then the estimation of n is based on the posterior distribution. Since prior probability distributions are assumed, more unknown parameters become involved. But in [2], the issue of how to select the values of the parameters has not been addressed. In this research memorandum, some guidelines for selecting the values of the parameters in the prior probability distributions are recommended. Recurrence formulas for calculating the posterior probabilities are also derived.

The method of estimating n from a Bayesian viewpoint is outlined in the next section. Suggestions for choosing values of unknown parameters in the prior probability distributions are discussed. Several examples are also presented in this section. Remarks and directions of future research are presented in the last section. Characteristics of beta distributions and an interactive computer program are included in appendixes.

THE METHOD

Let $X_1, X_2, \ldots, X_r$ be a random sample of size r from a binomial distribution $b(n, p)$. Given the observations $x_1, x_2, \ldots, x_r$, the objective is to estimate n. When $r = 1$ and p is known, the method of constructing confidence intervals for n has been presented in [1]. For r being a positive integer larger than one, Draper and Guttman [2]

proposed a Bayesian approach for estimating n. Adopting their notations, the likelihood can be written as:

$$L(n, p\ x) \propto p^t (1-p)^{rn-t} \prod_{i=1}^{r} \frac{n!}{(n-x_i)!} \ , \tag{1}$$

where $x = (x_1, x_2, \ldots, x_r)'$ is a column vector of positive integers and $t = \sum_{i=1}^{r} x_i$ is the total number of successes in the r observations. The case of p being known will be discussed first; then, the case of p being unknown will be discussed.

p is Known

When p is known, let $h(n)$ denote the prior distribution of n. Without further knowledge of n, the discrete uniform distribution provides a reasonable form for $h(n)$.

$$h(n) = \begin{cases} \frac{1}{N} & \text{for } 1 \le n \le N \\ 0 & \text{elsewhere} \end{cases} \ , \tag{2}$$

where N is a predetermined large integer. The posterior distribution for n is given by

$$p\ (n \mid x, p) \propto (1-p)^{rn}\ h(n) \prod_{i=1}^{r} \frac{n!}{(n-x_i)!} \ . \tag{3}$$

The domain of $p(n \mid x, p)$ is the set of n such that $n = x_{max}, x_{max}+1, x_{max}+2, \ldots, N$, where

$$x_{max} = max\ \{x_1, x_2, \ldots, x_r\} \ .$$

The mode of the posterior distribution $p(n \mid x, p)$ given in expression 3, denoted by $\hat{n}$, provides an estimate of n. $\hat{n}$, therefore, is the integer satisfying the following inequalities:

$$p\ (\hat{n}-1 \mid x, p) \le p\ (\hat{n} \mid x, p) \ ,$$

and

$$p\ (\hat{n}+1 \mid x, p) \le p\ (\hat{n} \mid x, p) \ .$$

Or alternatively, $\hat{n}$ is the solution of the simultaneous inequalities:

$$\prod_{i=1}^{r} (\hat{n} - x_i) \le [\hat{n}\ (1-p)]^r \ ,$$

and

$$[(\hat{n}+1)(1-p)]^r \le \prod_{i=1}^{r} (\hat{n}+1-x_i) \ .$$

It has been pointed out by Feldman and Fox [3] that $\hat{n}$ is also the maximum likelihood estimator for n.

In addition to providing an estimate for n, the posterior distribution could also cast some light on the precision of the estimate. A closed form of the estimator for n may not seem feasible. But a numerical solution can be obtained by using the following recurrence formula. For $n = x_{max} + j, j = 0, 1, 2, \ldots, (N - x_{max})$,

$$p(n \mid x, p) = p(x_{max} + j \mid x, p) = C \cdot Q_j \ , \tag{4}$$

where

$$Q_j = \begin{cases} 1 & \text{if } j = 0 \\ Q_{j-1} (1-p)^r \dfrac{(x_{max}+j)^r}{\prod_{i=1}^{r} (x_{max}+j-x_i)} & \text{, otherwise} \ . \end{cases} \tag{5}$$

Thus, the normalizing constant C in expression 4 is the reciprocal of the sum of the Q_j's, i.e.,

$$C = 1 / \sum_{j=0}^{N-x_{max}} Q_j \ . \tag{6}$$

As far as the point estimation is concerned, an estimate for n can be obtained irrespective of the predetermined integer N. If a confidence interval with a specified confidence coefficient γ is desired, then the value of N is needed. A 100 γ-percent confidence interval for n is given by

$$[x_{max} + \ell, x_{max} + u] \ , \tag{7}$$

where ℓ and u are integers such that

$$\sum_{j=1}^{u} p(x_{max} + j \mid x, p) \doteq \gamma \ . \tag{8}$$

Since ℓ and u in expression 7 are the integers satisfying the condition 8, they are chosen such that the summation on the left-hand side of 8 is approximately equal to γ and $1-\gamma$ is roughly equally divided to the two tails. Therefore, a 100 γ-percent confidence interval for n may not be unique.

To determine a suitable value of N for computing a confidence interval for n, one may adopt the scheme given below. Let

$$S_j = \sum_{i=0}^{j} Q_i \tag{9}$$

be the jth partial sum of the sequence $Q_0, Q_1, Q_2, \ldots$. For a given $\delta > 0$, j is defined to be the smallest integer such that

$$Q_j / S_j < \delta \, . \tag{10}$$

Therefore, the required integer N is equal to $x_{max} + j - 1$. The criterion stated in the inequality 10 suggests that the posterior probabilities, beyond the value of N, will not contribute significantly.

Example 1: r = 1

Suppose p is known to be 0.2. The only success count shows that ten successes have been detected, i.e., $x = 10$. Hence, x_{max} is also equal to 10. An estimate for n could be either 49 or 50. Similarly, 48 and 51 are also likely values for the estimate. Since $x = 10$ and $p = 0.2$, $x/p = 10/0.2 = 50$. The possible estimates are all close to 50.

Using the criterion 10 for $\delta = 0.005$, N is found to be 81. A 95-percent confidence interval for n is [30, 77]. As mentioned earlier, the confidence coefficient, 95 percent, is only an approximation.

Example 2: r = 4

Assume again $p = 0.2$. Four success counts are available: 4, 8, 12, and 8. Hence, $x_{max} = 12$. An estimate for n is eaual to 40. Other possible values are 39 and 41.

For $\delta = 0.005$, a 95-percent confidence interval can be chosen to be [31, 54] or [26, 51]. Both intervals give a confidence coefficient only to approximately 95 percent.

p is Unknown

When p is unknown, assume that n and p are independent. Let n have the same prior probability distribution as stated in expression 2. Suppose that the prior probability distribution for p is in the form of a beta distribution with parameters v_1 and v_2. Let $k(p)$ denote the prior probability distribution of p. Thus,

$$k(p) \propto p^{v_1-1} (1-p)^{v_2-1}, \; 0 \le p \le 1 \; . \tag{11}$$

Differentiate $k(p)$ in expression 11 with respect to p and let the first derivative of $k(p)$ equal zero. Then equation 12 represents the relationship among the maximum value and the two unknown parameters, v_1 and v_2, in the beta prior distribution:

$$(1-p)\, v_1 - p v_2 - 1 + 2p = 0 \; , \tag{12}$$

with an initial estimate of p; v_1 and v_2 can be solved through equation 12. However, the solution for v_1 and v_2 is not unique. If the initial estimate of p was obtained with high certainty, then the parameters v_1 and v_2 should be chosen with larger values, e.g., $v_1 = 10$ or 20. Otherwise, use smaller values; for example, v_1 may be 2 or 3. This recommendation is based on the characteristics of beta distributions. Beta distributions are further discussed in appendix A.

The joint posterior distribution is given by

$$p(n, p \mid x) \propto p^{\,t+v_1-1} (1-p)^{\,rn-t+v_2-1} \, h(n) \prod_{i=1}^{r} \frac{n!}{(n-x_1)!} \; . \tag{13}$$

The marginal distribution of n can be obtained by integrating expression 13 with respect to p from 0 to 1. Therefore,

$$p(n \mid x) \propto \frac{(rn - t + v_2 - 1)!}{(rn + v_1 + v_2 - 1)!} \prod_{i=1}^{r} \frac{n!}{(n-x_i)!} \; , \text{ for } x_{max} \le n \le N \; . \tag{14}$$

Again, the mode n of expression 14 would provide an estimate of n. Similarly, if $n = x_{max} + j$, for $j = 0, 1, 2, \ldots, N - x_{max}$,

$$p(n \mid x) = p(x_{max} + j \mid x) = C \cdot Q_j \; , \tag{15}$$

where

$$Q_j = \begin{cases} 1 & \text{if } j=0 \\ Q_{j-1} \prod_{1=0}^{r-1} \frac{[r x_{max} - t + v_2 + (j-1) r + i]}{[r x_{max} + v_1 + v_2 + (j-1) r + i]} \cdot \\ \frac{(x_{max} + j)^r}{\prod_{i=1}^{r} (x_{max} + j - x_i)} & \text{, otherwise} \end{cases} \tag{16}$$

is a recurrence formula for calculating Q_j. The normalizing constant C in expression 15 can be computed in exactly the same manner as before.

Example 3: r = 1

Suppose p is unknown. An initial p is found to be 0.2. Assume that the only success count gives ten successes. If δ in criterion 10 is chosen to be 0.005, then the estimates at various levels of certainty are presented in table 1.

TABLE 1

ESTIMATES OF n WHEN $r = 1$ AND p IS UNKNOWN

	n	N	95% confidence interval
$v_1 = 2, v_2 = 5$	29 or 30	106	[17, 102]
$v_1 = 5, v_2 = 17$	41 or 42	100	[23, 96]
$v_1 = 10, v_2 = 37$	45 or 46	93	[26, 89]
$v_1 = 20, v_2 = 77$	47 or 48	88	[27, 83]

When the initial estimate of p is made with high certainty, such as $v_1 = 20$ and $v_2 = 77$, the point estimate for n is almost identical to the result given in example 1, in which p is known. However, with p unknown, confidence intervals are not as tigh.

Example 4: $r = 4$

Consider the set of data given in example 2. Assume that an initial estimate for p is 0.2 and $\delta = 0.005$. The results are summarized in table 2.

TABLE 2

ESTIMATES OF n WHEN $r = 4$ AND p IS UNKNOWN

	n	N	95% confidence interval
$v_1 = 2$, $v_2 = 5$	29	103	[17, 97]
$v_1 = 5$, $v_2 = 17$	35	85	[22, 81]
$v_1 = 10$, $v_2 = 37$	38	74	[24, 69]
$v_1 = 20$, $v_2 = 77$	39	67	[26, 62]

CONCLUSIONS AND REMARKS

In this research, the procedure for estimating n in a binomial distribution has been developed if there are r success counts available. Both cases of p being known and cases of p being unknown are discussed. The approach is adopted from a Bayesian point of view, which was first proposed by Draper and Guttman in [2]. Prior probability distribution for n is assumed to be the discrete uniform distribution. When p is unknown, p is assumed to have a beta prior probability distribution. The estimator for n is then the mode of the posterior distribution.

To put the estimation procedure into practice, suggestions of how to select values of unknown quantities in the prior probability distribution are provided. An interactive computer program, written in FORTRAN language, is included in appendix B.

The interval estimation for n is also discussed in this paper. As pointed out in the paper, confidence coefficients are not exact because n under consideration is discrete. Furthermore, confidence intervals are not unique.

Future study may include a simulation to verify whether the proposed procedure for constructing confidence intervals for the specified confidence coefficients has actually been achieved. Another topic worth further investigation is the assessment of the error of the estimators. The bootstrap method may be a useful tool for the investigation.

As mentioned by several authors before, e.g., Carroll and Lombard [4], estimators for n in a binomial distribution are usually unstable. Therefore, the value of p used in the estimation procedure must be selected with special caution.

REFERENCES

[1] CNA Research Memorandum 86-265, *Confidence Intervals for Parameter* n *in a Binomial Distribution*, by Victor K. T. Tang and Ronald B. Sindler, Jan 1987 (27860265)[1]

[2] N. Draper and I. Guttman. "Bayesian Estimation of the Binomial Parameter." *Technometrics* 13, No. 3 (1971): 667-673

[3] D. Feldman and M. Fox. "Estimation of the Parameter n in the Binomial Distribution." *Journal of the American Statistical Association* 63 (1968): 150-158

[4] R. U. Carroll and F. Lombard. "A Note on N Estimators for the Binomial Distribution." *Journal of the American Statistical Association* 80 (1985): 423-426

1. The number in parentheses is a CNA internal control number.

APPENDIX A

CHARACTERISTICS OF BETA DISTRIBUTIONS

APPENDIX A

CHARACTERISTICS OF BETA DISTRIBUTIONS

Let Y be a random variable having a beta distribution with parameters v_1 and v_2. The probability density function of Y is given by

$$f(y) = \begin{cases} \dfrac{\Gamma(v_1 + v_2)}{\Gamma(v_1) \cdot \Gamma(v_2)} \, y^{v_1-1} \, (1-y)^{v_2-1}, & 0 \le y \le 1 \,, \\ 0 & , \text{elsewhere} \end{cases} \tag{A-1}$$

where v_1 and v_2 are positive and $\Gamma(.)$ denotes the gamma function. The shape of the density of $f(y)$ given in equation A-1 depends on the values of v_1 and v_2.

Case 1: $v_1 = v_2 = 1$

$f(y)$ in (A-1) becomes

$$f(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases} \,, \tag{A-2}$$

which is the density of the uniform distribution in the unit interval. The graph of (A-2) is shown in figure A-1.

Case 2: $v_1 = v_2 > 1$

Equation A-1 becomes

$$f(y) = \begin{cases} \dfrac{\Gamma(2v_1)}{[\Gamma(v_1)]^2} \, [y\,(1-y)]^{v_1-1}, & 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases} \,. \tag{A-3}$$

The function $f(y)$ in expression A-3 is symmetric with respect to the line $y = 1/2$. The mode occurs at $y = 1/2$. The mode is

$$f(1/2) = \frac{\Gamma(2v_1)}{[\Gamma(v_1)]^2} \left(\frac{1}{4}\right)^{v_1-1} .$$

Graphs of (A-3), with various values of v_1, are shown in figure A-2.

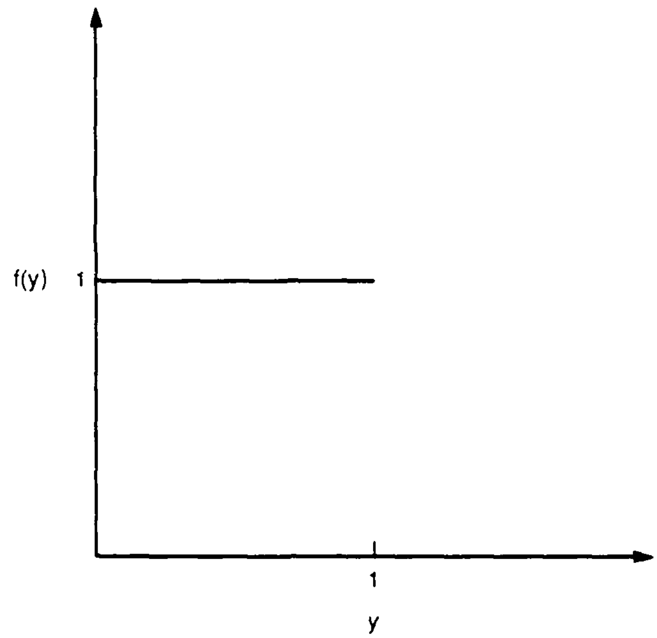

FIG. A-1: $\nu_1 = \nu_2 = 1$

Case 3: $\nu_2 > \nu_1 > 1$

Equation A-1 is not symmetric. The mode occurs at

$$y = (\nu_1 - 1) / [(\nu_1 - 1) + (\nu_2 - 1)] \ .$$

In this case, the mode y is less than one-half. The density shows a positive skewness; in other words, the graph has a long tail to the right. Graphs are presented in figure A-3 for various values of ν and ν with the mode equal to 0.2.

Case 4: $\nu_1 > \nu_2 > 1$

Case 4 is the reverse of case 3. The mode is larger than one-half. The density shows a negative skewness; therefore, it has a long tail to the left. Graphs of beta distributions in this are the symmetric images with respect to the line $y = 1/2$ of those given in case 3 and shown in figure A-3.

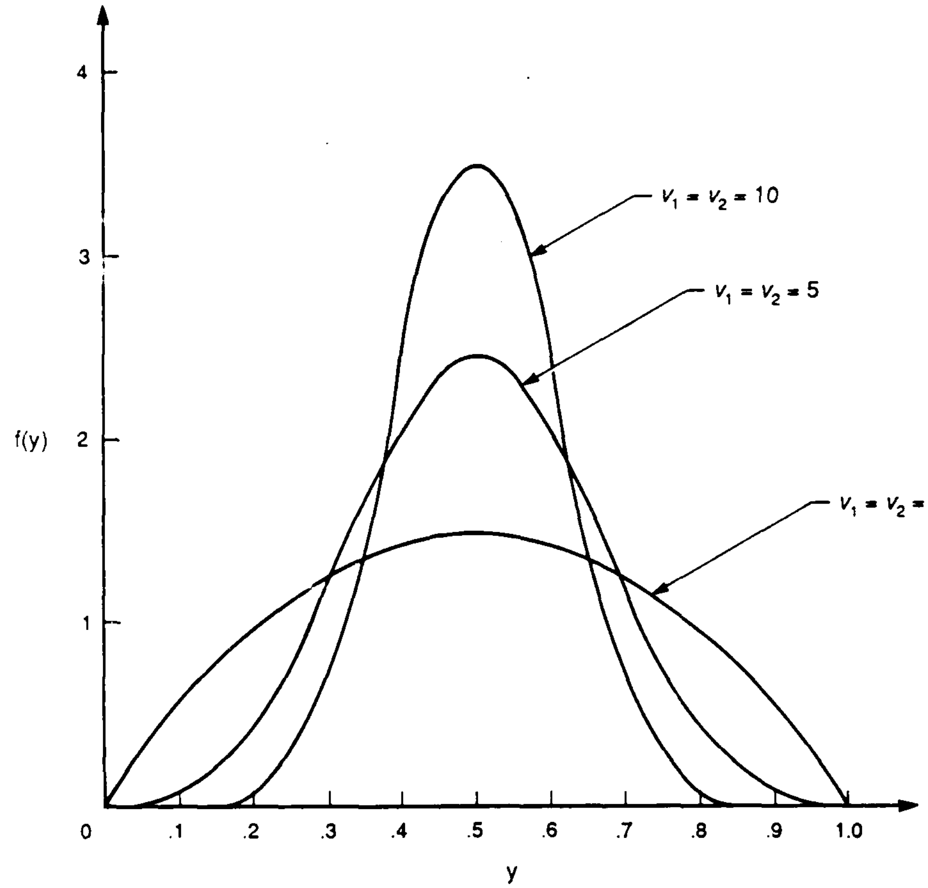

FIG. A-2: $\nu_1 = \nu_2 > 1$

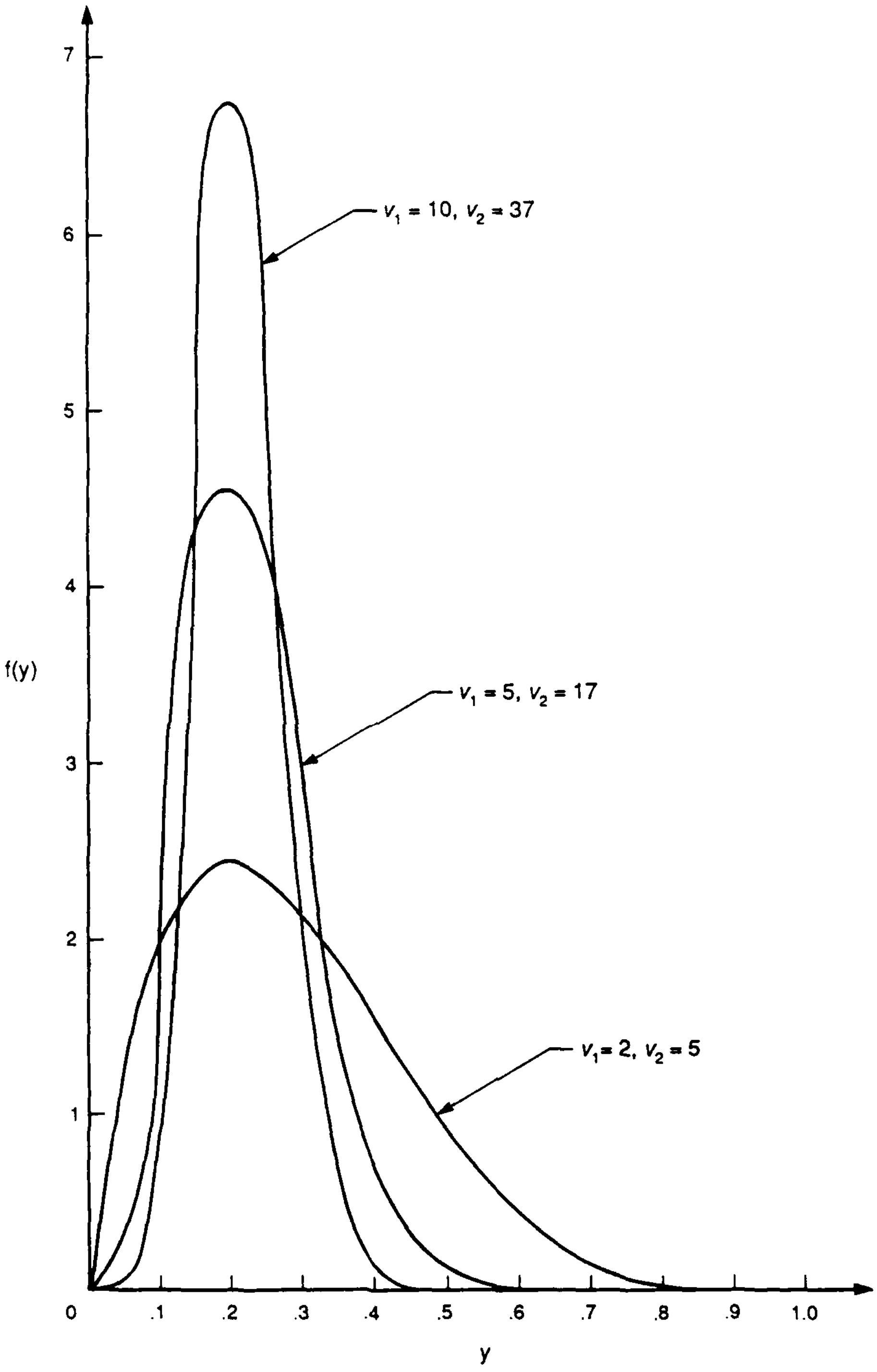

FIG. A-3: $v_2 > v_1 > 1$

APPENDIX B

AN INTERACTIVE COMPUTER PROGRAM

```
C     BAYESIAN ESTIMATION OF n IN A BINOMIAL DISTRIBUTION
C
C     PROGRAM INPUT  1). OBSERVATIONS FOR RANDOM VARIABLE X(J)
C                    2). PROBABILITY OF SUCCESS (PRSUCC)
C                    3). SHAPE PARAMETER FOR PRIOR BETA DISTRIBUTION (V1)
C     PROGRAM OUTPUT 1). POSTERIOR PROBABILITIES
C                    2). NUMBER OF TRIALS (n)
C
C     VARIABLE DEFINITIONS
C
C     GAMMA     - THE NUMBER OF SUCCESS COUNTS (NUMBER OF X(J)'s)
C     QSUM      - SUM OF Q
C     POSTPROB  - COMPUTED POSTERIOR PROBABILITY
C     PRSUCC    - PROBABILITY OF SUCCESS (BINOMIAL)
C     TSUM      - TOTAL NUMBER OF SUCCESSES (SUM OF X(J)'s)
C     V1        - SHAPE PARAMETER OF BETA PRIOR DISTRIBUTION
C     V2        - SHAPE PARAMETER OF BETA PRIOR DISTRIBUTION
C     XJ        - OBSERVATION OF RANDOM VARIABLE X(J)
C                    (NUMBER OF OBJECTS DETECTED)
C     XJMAX     - MAXIMUM VALUE OF RANDOM VARIABLE X(J)
C     MODE      - MODE OF PROBABILITY DISTRIBUTUIN. THIS IS
C                    THE n VALUE THAT HAS THE MAXIMUM PROBABILITY
C     LSUM      - SUM OF PROBABILITIES IN LOWER TAIL OF DISTRIBUTION
C                    FOR A 95% CONFIDENCE INTERVAL
C     USUM      - SUM OF PROBABILITIES IN UPPER TAIL OF DISTRIBUTION
C                    FOR A 95% CONFIDENCE INTERVAL
C     LBOUND    - LOWER BOUND OF CONFIDENCE INTERVAL FOR n
C     UBOUND    - UPPER BOUND OF CONFIDENCE INTERVAL FOR n
C
C     DATA DECLARATION/INITIALIZATION
C
      INTEGER XJ(10), XJMAX, XVALUE, GAMMA, TSUM, N(1000), UBOUND
      DOUBLE PRECISION Q(1000), QSUM, POSTPROB(1000), MODE, LSUM, USUM
      CHARACTER * 1 KE
      DELTA  = .005
      XJMAX  = 0
      TSUM   = 0
      QSUM   = 0.0
      DATA N  /1000*0/
      DATA XJ /10*0/
C
C     ACCEPT XJ's FROM INPUT DEVICE AND COMPUTE GAMMA, TSUM, XJMAX
C
      DO 20 I=1,10
 10      PRINT *, ' '
         PRINT *, 'ENTER VALUE FOR THE NUMBER OF OBSERVATIONS IN SAMPLE
     & X(j) (UP TO 10 SAMPLES) '
         PRINT *, 'PRESS RETURN WHEN ALL SAMPLES HAVE BEEN ENTERED '
         PRINT *, ' '
         ACCEPT 15, XVALUE
 15      FORMAT (I80)
         IF (XVALUE .EQ. 0) THEN
            IF (I .GT. 1) THEN
               GOTO 35
            ELSE
               PRINT *, ' '
               PRINT *, 'MUST ENTER AT LEAST ONE VALUE '
               GOTO 10
            ENDIF
         ENDIF
         GAMMA  = I
```

```
         XJ(I)   = XVALUE
         TSUM    = TSUM + XVALUE
C
C        DETERMINE MAXIMUM SAMPLE VALUE
C
         IF (XVALUE .GE. XJMAX) XJMAX = XVALUE
C
   20 CONTINUE
C
C     ACCEPT PROBABILITY OF DETECTION FROM INPUT DEVICE AND EDIT VALUE
C
   35 PRINT *, ' '
      PRINT *, 'ENTER PROBABILITY OF SUCCESS '
      PRINT *, '(MUST BE .1, .2 .3, .4, OR .5) '
      PRINT *, ' '
      ACCEPT 40, PRSUCC
   40 FORMAT (F80.2)
      IF ((PRSUCC .EQ. .1) .OR.
     &    (PRSUCC .EQ. .2) .OR.
     &    (PRSUCC .EQ. .3) .OR.
     &    (PRSUCC .EQ. .4) .OR.
     &    (PRSUCC .EQ. .5)) THEN
         GOTO 41
      ELSE
         PRINT *, ' '
         PRINT *, 'PROBABILITY OF SUCCESS MUST BE .1, .2, .3, OR .4'
      GOTO 35
      ENDIF
   41 PRINT *, ' '
      PRINT *, 'IS PROBABILITY OF SUCCESS KNOWN (K) OR ESTIMATED (E)'
      PRINT *, 'ENTER K OR E'
      PRINT *, ' '
      ACCEPT 42, KE
   42 FORMAT (A1)
      IF ((KE .EQ. 'E') .OR. (KE .EQ. 'e')) THEN
         GOTO 45
      ELSEIF ((KE .EQ. 'K') .OR. (KE .EQ. 'k')) THEN
         GOTO 500
      ELSE
         PRINT *, ' '
         PRINT *, 'MUST ENTER K OR E '
         GOTO 41
      ENDIF
C
C     COMPUTE PARAMETER V2 FOR BETA PRIOR DISTRIBUTION. A VALUE FOR
C     SHAPE PARAMETER V1 (FLOATING POINT VALUES 2 THROUGH 10) IS ACCEPTED
C     FROM THE INPUT DEVICE. A LARGER VALUE OF V1 INDICATES GREATER
C     CERTAINTY FOR THE SELECTED PROBABILITY OF SUCCESS. V2 IS
C     THEN COMPUTED AS A FUNCTION OF V1 BASED ON THE FOLLOWING
C     EQUATION. (V1 -1)(1-p) - (V2-1)(p) = 0, WHERE p IS THE
C     PROBABILITY OF SUCCESS.
C
   45 PRINT *, ' '
      PRINT *, 'ENTER SHAPE PARAMETER FOR PRIOR BETA DISTRIBUTION'
      PRINT *, '(PARAMETER MUST TAKE ON FLOATING POINT VALUES BETWEEN 2.
     &0 AND 10.0) '
      PRINT *, ' '
      PRINT *, 'LARGER VALUES INDICATE GREATER CERTAINTY FOR THE SELECT
     &ED '
      PRINT *, 'PROBABILITY OF SUCCESS. '
      PRINT *, ' '
```

```
      ACCEPT 50, V1
   50 FORMAT (F80.2)
      IF (V1 .GE. 2.0 .AND. V1 .LE. 10) THEN
         CONTINUE
      ELSE
         PRINT *, ' '
         PRINT *, 'SHAPE PARAMETER MUST TAKE ON VALUES BETWEEN 2.0 AND 1
     &0.0'
         GOTO 45
      ENDIF
      IF (PRSUCC .EQ. 0.1)     THEN
         V2 =  (9.0 * V1) - 8.0
      ELSEIF (PRSUCC .EQ. 0.2) THEN
         V2 =  (4.0 * V1) - 3.0
      ELSEIF (PRSUCC .EQ. 0.3) THEN
         V2 = ((7.0 * V1) - 4.0) / 3.0
      ELSE
C ***    PRSUCC = 0.4
         V2 = ((3.0 * V1) - 1.0) / 2.0
      ENDIF
C
C     COMPUTE ESTIMATED PROBABILITY CASE
C
C     COMPUTE Q(I) AND QSUM. COMPUTATIONS ARE TERMINATED
C     WHEN Q(I)/QSUM IS LESS THAN OR EQUAL TO DELTA.
C
C     Q(I) IS THE PRODUCT OF THE FOLLOWING THREE FACTORS
C
C     1). Q(I-1)
C
C     2). THE PRODUCT FROM     (GAMMA*XJMAX-TSUM+V2+((I-1)*GAMMA)+J)
C         J=0 TO GAMMA-1       (GAMMA*XJMAX+V1+V2+((I-1)*GAMMA)+J)
C
C     3). ((XJMAX+I)**GAMMA) / THE PRODUCT FROM (XJMAX-XJ(J)+I)
C                              J=1 TO GAMMA
C
C     LET CONSTANT A = GAMMA*XJMAX-TSUM+V2
C     LET CONSTANT B = GAMMA*XJMAX+V1+V2
C
      A = FLOAT(GAMMA)*FLOAT(XJMAX)-FLOAT(TSUM)+V2
      B = FLOAT(GAMMA)*FLOAT(XJMAX)+V1+V2
      Q(1) = 1.0
      DO 200 I=1,9999
         ISTORE = I+1
         N(I) = XJMAX+I-1
         IF (N(I) .EQ. 129) GOTO 205
C
C        COMPUTE FACTOR 1
C
         IF (I .GT. 1) THEN
            FACTR1 = Q(I)
         ELSE
            FACTR1 = 1.0
         ENDIF
C
C        COMPUTE FACTOR 2
C
         FACTR2 = 1.0
         DO 105 J=0,GAMMA-1
            FACTR2 = FACTR2 * (A+((I-1)*GAMMA)+J)/(B+((I-1)*GAMMA)+J)
  105    CONTINUE
```

```
C
C        COMPUTE FACTOR 3
C
         PRODCT = 1.0
         DO 110 J=1,GAMMA
            PRODCT = PRODCT * (FLOAT(XJMAX)-FLOAT(XJ(J))-FLOAT(I))
  110    CONTINUE
         FACTR3 = (FLOAT(XJMAX-I)**GAMMA)/PRODCT
C
C        COMPUTE Q(I) AND QSUM
C
         Q(I+1) = FACTR1*FACTR2*FACTR3
         QSUM = QSUM + Q(I)
C
C        TEST TERMINATION CRITERIA
C
         IF (Q(I)/QSUM .GT. DELTA) THEN
            CONTINUE
         ELSE
            GOTO 205
         ENDIF
  200 CONTINUE
C
C     COMPUTE POSTERIOR PROBABILITY DISTRIBUTION.
C
  205 DO 300 I=1,ISTORE
         POSTPROB(I) = Q(I)/QSUM
  300 CONTINUE
      GOTO 702
C
C     COMPUTE KNOWN PROBABILITY CASE
C
C     COMPUTE Q(I) AND QSUM. COMPUTATIONS ARE TERMINATED
C     WHEN Q(I)/QSUM IS LESS THAN OR EQUAL TO DELTA.
C
C     Q(I) IS THE PRODUCT OF THE FOLLOWING THREE FACTORS
C
C     1). Q(I-1)
C
C     2). (1-P)**GAMMA
C
C     3). ((XJMAX-I)**GAMMA) / THE PRODUCT FROM (XJMAX-XJ(J)+I)
C                              J=1 TO GAMMA
  500 Q(1) = 1.0
      DO 600 I=1,9999
         ISTORE = I+1
         N(I) = XJMAX-I-1
         IF (N(I) .EQ. 120) GOTO 605
C
C        COMPUTE FACTOR 1
C
         IF (I .GT. 1) THEN
            FACTR1 = Q(I)
         ELSE
            FACTR1 = 1.0
         ENDIF
C
C        COMPUTE FACTOR 2
C
         FACTR2 = (1.0 - PRSUCC)**GAMMA
C
```

```
C        COMPUTE FACTOR 3
C
         PRODCT = 1.0
         DO 510 J=1,GAMMA
            PRODCT = PRODCT * (FLOAT(XJMAX)-FLOAT(XJ(J))+FLOAT(I))
  510    CONTINUE
         FACTR3 = (FLOAT(XJMAX+I)**GAMMA)/PRODCT
C
C        COMPUTE Q(I) AND QSUM
C
         Q(I+1) = FACTR1*FACTR2*FACTR3
         QSUM = QSUM + Q(I)
C
C        TEST TERMINATION CRITERIA
C
         IF (Q(I)/QSUM .GT. DELTA) THEN
            CONTINUE
         ELSE
            GOTO 605
         ENDIF
  600 CONTINUE
C
C     COMPUTE POSTERIOR PROBABILITY DISTRIBUTION.
C
  605 DO 610 I=1,ISTORE
         POSTPROB(I) = Q(I)/QSUM
  610 CONTINUE
C
C
C     GENERATE AND DISPLAY REPORT ON CRT
C
  702 WRITE  (6,705)
  705 FORMAT (1H0,'BAYESIAN ESTIMATION OF n IN A BINOMIAL DISTRIBUTI
     &ON')
      WRITE  (6,710) PRSUCC
  710 FORMAT (1H0,'PROBABILITY OF SUCCESS (ESTIMATE) = ',F3.1)
      IF ((KE .EQ. 'E') .OR. (KE .EQ. 'e')) THEN
         WRITE  (6,715) V1, V2
  715    FORMAT (1H ,'SHAPE PARAMETERS FOR PRIOR BETA V1 = ',F5.2,
     &           ' V2 = ',F5.2,/)
      ENDIF
      DO 730 I=1,10
         IF (XJ(I) .EQ. 0)     THEN
            GOTO 730
         ELSEIF (I .LE. 9) THEN
            WRITE  (6,720) I, XJ(I)
  720       FORMAT (1H ,'NUMBER OF OBSERVATIONS IN SAMPLE X(',I1,
     &              ')  = ',I2)
         ELSE
            WRITE  (6,725) I, XJ(I)
  725       FORMAT (1H ,'NUMBER OF OBSERVATIONS IN SAMPLE X(',I2,
     &              ') = ',I2)
         ENDIF
  730 CONTINUE
      DO 733 I=1,9999
         IF (POSTPROB(I) .GT. POSTPROB(I+1)) THEN
            IMODE = I
            MODE  = POSTPROB(I)
            GOTO 734
         ENDIF
  733 CONTINUE
```

```
  734 WRITE  (6,735)
  735 FORMAT (1H0,'    n      p(n:x,p)')
      WRITE  (6,740)
  740 FORMAT (1H ,'    --------------' )
      DO 750 I=IMODE-5,IMODE+5
         WRITE  (6,745) N(I), POSTPROB(I)
  745    FORMAT (1H ,2X,I3,4X,F7.6)
  750 CONTINUE
C
C     GENERATE 95% CONFIDENCE INTERVAL FOR n
C
      LSUM = 0.0
      DO 770 I=1,9999
         LSUM = LSUM + POSTPROB(I)
         IF (LSUM .LE. 0.025) THEN
            LBOUND = N(I+1)
            GOTO 770
         ELSE
            LSUM = LSUM - POSTPROB(I)
            GOTO 780
         ENDIF
  770 CONTINUE
C
C     N = ISTORE-2
C
  780 USUM = LSUM
      DO 790 I=ISTORE-2,1,-1
         USUM = USUM + POSTPROB(I)
         IF (USUM .LE. 0.05) THEN
            UBOUND = N(I-1)
            GOTO 790
         ELSE
            GOTO 800
         ENDIF
  790 CONTINUE
  800 WRITE  (6,810) N(IMODE)
  810 FORMAT (1H ,/,' MODE OF DISTRIBUTION IS n = ',I3)
      WRITE  (6,820) LBOUND
  820 FORMAT (1H0,'LOWER BOUND OF 95% CONFIDENCE INTERVAL = ',I3)
      WRITE  (6,830) UBOUND
  830 FORMAT (1H ,'UPPER BOUND OF 95% CONFIDENCE INTERVAL = ',I3)
C
      PRINT *, ' '
      STOP
      END
```